Classification of online handwriting time series for Parkinson's Disease diagnosis using Deep Learning

Paul Lerner[1 [0000−0002−0882−8684]] and
Laurence Likforman-Sulem[1 [0000−0001−9096−7239]]

Télécom ParisTech / Université Paris-Saclay. 46 Rue Barrault, 75013 Paris, France
1.paul.lerner@gmail.com
likforman@telecom-paristech.fr

Abstract. Parkinson's Disease (PD) is the second most common neurodegenerative disease. Several of its symptoms are observable through handwriting analysis at an early stage of the disease. A few authors have already used Machine Learning techniques in order to classify PD and Healthy Control (HC) subjects based on hand-crafted features extracted from the data. We decided to focus on Deep Learning models which are able to learn features from the raw data. We consider here the *PaHaW* database which consists of seven handwriting tasks represented as time series.
We present two different model architectures : a Convolutional Neural Network (1d-CNN) and a Hierarchical CNN-Recurrent Neural Network (CNN-RNN). The latter provides promising results for end-to-end learning on a small dataset like ours. We believe it is because of its hierarchical structure which augments the dataset.
The CNN-RNN outperforms the state-of-the-art on some tasks but it is not suited for all. Thus, after combining the prediction over all tasks, the CNN-RNN still falls below [1] who use hand-crafted features and [4] who use transfer learning. For this reason we plan to explore transfer learning in future works.

Keywords: Parkinson's Disease · Deep Learning · end-to-end learning · Handwriting features · Time series · Online Handwriting

1 Motivation

Parkinson's Disease (PD) is the second most common neurodegenerative disease. Its diagnosis is considered to be difficult as there is a large number of symptoms which are shared among other diseases (see, e.g. [2]). However, the diagnosis is usually assessed by a neurologist after a physical examination as SPECT and CT scans are costly, invasive, and usually effective when the disease has already progressed to a mature stage [4]. In the description of the disease made by James Parkinson in 1817, writing deficits precede walking deficits. Since then, handwriting has been used to assess PD because symptoms such as *tremor* or *micrographia* may be visible through a traditional paper-and-pencil examination. However, for a few years, researchers have been arguing that PD *dysgraphia* is larger than *micrographia* and that it is observable through *kinematic* analysis [3]. The goal for the researchers is to provide for a Clinical Decision Support System which would confirm or question the neurologist' diagnosis based on a cheap and non-invasive handwriting exam.

Several authors have already used Machine Learning techniques in order to classify PD and HC based on hand-crafted features extracted from the data. However PD's handwriting impairments result from numerous symptoms [7], thus it is very difficult to decide which measure to extract and some inconsistent results have been obtained depending on the handwriting task [1]. Therefore we decided to focus on Deep Learning models which are able to learn features from the raw data. [6] had already performed end-to-end learning of handwriting features for PD diagnosis but on another database where HC are in average 14 years younger than PD. As numerous symptoms of PD are shared with normal aging [5], we prefer to work on the *PaHaW* database.

We compare our work to [1] who use a Support Vector Machine (SVM) classifier over hand-crafted features and [4] who use a 2d-CNN pre-trained on *ImageNet* over static images. Both use the *PaHaW* database which is briefly described in the next section.

2 Methods and Results

We use the publicly-available *PaHaW* database which comprises 72 subjects (36 PDs and 36 HCs) who all performed seven handwriting tasks : an Archimedean spiral, five *l*, five *le*, five *les*, three Czech words and one Czech sentence [1]. Each task was recorded through a tablet as a multivariate time-serie of seven measures : the x and y spatial coordinates, a time stamp, the button status which indicates whether the stroke is in-air or on-paper, the pressure and, the tilt and elevation which are angles that the pen makes with z and y axis, respectively.

We propose two different architectures: a 1d-CNN which takes the whole task as an input and a CNN-RNN which takes a list of strokes as an input (see Fig. 1).

The CNN-RNN reaches 73% classification accuracy after majority voting over all tasks and 10-fold cross-validation, thus outperforming the 1d-CNN which only reaches 63% classification accuracy. We believe this is because the CNN-RNN has a Hierarchical structure which augments the dataset. It can also explained because of the distinction made between in-air and on-paper strokes (see Fig. 1). Moreover the CNN-RNN outperforms [1] and [4] on the *l*, *le* and *les* tasks. We believe it is because in those tasks, the number of strokes is highly less variable between subjects than in other tasks where our models performs rather poorly. Thus, after majority voting on all tasks, the CNN-RNN still falls below [1] and [4] who achieve 81% and 83% classification accuracy, respectively.

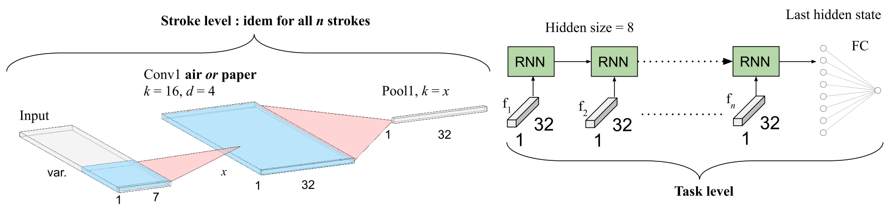

Fig. 1. CNN-RNN architecture. k stands for kernel size and d stands for dilation (*aka* à trous convolution). Note that $Conv1$ is either $Conv1_{air}$ or $Conv1_{paper}$ and gets fed in-air and on-paper strokes, respectively.

3 Discussion

The biggest challenge of this work was the size of the dataset (72 subjects) which inhibited the generalization capacity of the model : even when using a very small network compared to deep learning standards we still heavily overfit the training set. We expect a great improvement of the results when using a larger dataset. Augmenting the dataset by splitting data into strokes has given interesting results with the CNN-RNN, however classical data augmentation techniques did not provide encouraging results.

In future works we will explore transfer learning techniques, which have proven to be quite efficient in [4]. We plan to pre-train the model to tasks such as mode detection or auto-encoding on large online handwriting databases.

References

1. Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M.: Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. Artificial intelligence in Medicine **67**, 39–46 (2016)
2. Hughes, A.J., Daniel, S.E., BenShlomo, Y., Lees, A.J.: The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain **125**(4), 861–870 (04 2002)
3. Letanneux, A., Danna, J., Velay, J.L., Viallet, F., Pinto, S.: From micrographia to parkinson's disease dysgraphia. Movement Disorders **29**(12), 1467–1475 (2014)
4. Moetesum, M., Siddiqi, I., Vincent, N., Cloppet, F.: Assessing visual attributes of handwriting for prediction of neurological disorders - a case study on parkinson's disease. Pattern Recognition Letters **121**, 19–27 (2019)
5. Mortimer, J.A.: Human motor behavior and aging. Annals of the New York Academy of Sciences **515**(1), 54–66 (1988)
6. Pereira, C.R., Pereira, D.R., Rosa, G.H., Albuquerque, V.H., Weber, S.A., Hook, C., Papa, J.P.: Handwritten dynamics assessment through convolutional neural networks: An application to parkinson's disease identification. Artificial intelligence in medicine **87**, 67–77 (2018)
7. Phillips, J., Stelmach, G.E., Teasdale, N.: What can indices of handwriting quality tell us about parkinsonian handwriting? Human Movement Science **10**(2-3), 301–314 (1991)